OIPE

RAW SEQUENCE LISTING   DATE: 05/30/2001
PATENT APPLICATION: **US/09/724,296**   TIME: 15:37:00

Input Set : **A:\25-98a.app**
Output Set: **C:\CRF3\05302001\I724296.raw**

ENTERED

```
 3 <110> APPLICANT: Doetsch, Paul W.
 4        Avery, Angela M.
 5        Kaur, Balveen
 7 <120> TITLE OF INVENTION: Broad Specificity DNA Damage Endonuclease
 9 <130> FILE REFERENCE: 25-98A
11 <140> CURRENT APPLICATION NUMBER: US/09/724,296
12 <141> CURRENT FILING DATE: 2000-11-28
14 <150> PRIOR APPLICATION NUMBER: US/09/327,984
15 <151> PRIOR FILING DATE: 1999-06-08
17 <150> PRIOR APPLICATION NUMBER: US/60/088,521
18 <151> PRIOR FILING DATE: 1998-06-08
20 <150> PRIOR APPLICATION NUMBER: US/60/134,752
21 <151> PRIOR FILING DATE: 1999-05-18
23 <160> NUMBER OF SEQ ID NOS: 71
25 <170> SOFTWARE: PatentIn Ver. 2.0
27 <210> SEQ ID NO: 1
28 <211> LENGTH: 2492
29 <212> TYPE: DNA
30 <213> ORGANISM: Artificial Sequence
32 <220> FEATURE:
33 <223> OTHER INFORMATION: Description of Artificial Sequence: Coding
34        sequence for fusion protein of GST signal peptide
35        and the UVDE protein of Schizosaccharomyces pombe
37 <400> SEQUENCE: 1
38 atgaccaagt tacctatact aggttattgg aaaaattaag ggccttgtgc aacccactcg 60
39 acttcttttg gaatatcttg aagaaaaata tgaagagcat ttgtatgagc gcgatgaagg 120
40 tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc ttccttatta 180
41 tattgatggt gatgttaaat taacacagtc tatggccatc atacgttata tagctgacaa 240
42 gcacaacatg ttggttggtt gtccaaaaga gcgtgcagag atttcaatgc ttgaaggagc 300
43 ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact ttgaaactct 360
44 caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag atcgtttatg 420
45 tcataaaaca tatttaaatg ttgaccatgt aacccatcct gacttcatgt tgtatgacgc 480
46 tcttgatgtt gttttataca tggacccaat gtgcctggat gcgttcccaa aattagtttg 540
47 tttttaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat ccagcaagta 600
48 tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc atcctccaaa 660
49 atcggatcat ctggttccgc gtggatccat gcttaggcta ttgaaacgaa atattcaaat 720
50 ctctaaacgc attgttttca ccatattaaa acaaaaggca tttaaaggta atcatccttg 780
51 tgtaccgtcg gtttgtacca ttacttactc tcgttttcat tgtttacccg ataccttaa 840
52 aagtttactt ccaatgagct caaaaaccac actctcaatg ttaccgcaag ttaatatcgg 900
53 tgcgaattca ttctctgccg aaacaccagt cgacttaaaa aagaaaatg agactgagtt 960
54 agctaatatc agtggacctc acaaaaaaag tacttctacg tctacacgaa agagggcacg 1020
55 tagcagtaaa aagaaagcga cagattctgt ttccgataaa attgatgagt ctgttgcgtc 1080
56 ctatgattct tcaactcatc ttaggcgatc gtcgagatca aaaaaaccgg tcaactacaa 1140
57 ttcctcgtca gaatccgaat cggaggagca aattagtaaa gctactaaaa aagttaaaca 1200
58 aaaagaggaa gaggagtatg ttgaagaagt cgacgaaaag tctcttaaaa atgaaagtag 1260
59 ctctgacgag ttcgaaccgg ttgtgccgga acagttggaa actccaattt ctaaacgaag 1320
60 acggtctcgt tcttctgcaa aaaatttaga aaaagaatct acaatgaatc ttgatgatca 1380
```

Input Set : A:\25-98a.app
Output Set: C:\CRF3\05302001\I724296.raw

```
61 tgctccacga gagatgtttg attgtttgga caaacccata ccctggcgag gacgattggg 1440
62 gtatgcttgt ttgaatacta tttttaaggtc aatgaaggag agggtttttt gttcacgcac 1500
63 ctgccgaatt acaaccattc aacgtgatgg gctcgaaagt gtcaagcagc taggtacgca 1560
64 aaatgtttta gatttaatca aattggttga gtggaatcac aactttggca ttcacttcat 1620
65 gagagtgagt tctgatttat ttcctttcgc aagccatgca aagtatggat ataccccttga 1680
66 atttgcacaa tctcatctcg aggaggtggg caagctggca aataaatata atcatcgatt 1740
67 gactatgcat cctggtcagt acacccagat agcctctcca cgagaagtcg tagttgattc 1800
68 ggcaatacgt gatttggctt atcatgatga aattctcagt cgtatgaagt tgaatgaaca 1860
69 attaaataaa gacgctgttt taattattca ccttggtggt acctttgaag gaaaaaaga 1920
70 aacattggat aggtttcgta aaaattatca acgcttgtct gattcggtta aagctcgttt 1980
71 agttttagaa aacgatgatg tttcttggtc agttcaagat ttattacctt tatgccaaga 2040
72 acttaatatt cctctagttt tggattggca tcatcacaac atagtgccag gaacgcttcg 2100
73 tgaaggaagt ttagatttaa tgccattaat cccaactatt cgagaaacct ggacaagaaa 2160
74 gggaattaca cagaagcaac attactcaga atcggctgat ccaacggcga tttctgggat 2220
75 gaaacgacgt gctcactctg ataggtgtgtt tgactttcca ccgtgtgatc ctacaatgga 2280
76 tctaatgata gaagctaagg aaaaggaaca ggctgtattt gaattgtgta gacgttatga 2340
77 gttacaaaat ccaccatgtc ctcttgaaat tatggggcct gaatacgatc aaactcgaga 2400
78 tggatattat ccgcccggag ctgaaaagcg tttaactgca agaaaaaggc gtagtagaaa 2460
79 agaagaagta gaagaggatg aaaaataaaa at                                2492
81 <210> SEQ ID NO: 2
82 <211> LENGTH: 828
83 <212> TYPE: PRT
84 <213> ORGANISM: Artificial Sequence
86 <220> FEATURE:
87 <223> OTHER INFORMATION: Description of Artificial Sequence: Fusion protein
88       of GST leader peptide and Schizosaccharomyces
89       pombe UVDE
91 <400> SEQUENCE: 2
92 Met Thr Lys Leu Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val
93  1               5                  10                  15
95 Gln Pro Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu
96              20                  25                  30
98 His Leu Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe
99          35                  40                  45
101 Glu Leu Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp
102      50                  55                  60
104 Val Lys Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys
105  65                  70                  75                  80
107 His Asn Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met
108              85                  90                  95
110 Leu Glu Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala
111          100                 105                 110
113 Tyr Ser Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu
114          115                 120                 125
116 Pro Glu Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr
117      130                 135                 140
119 Leu Asn Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala
120 145                 150                 155                 160
122 Leu Asp Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro
```

5/30/01

```
123                 165                 170                 175
125 Lys Leu Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp
126                 180                 185                 190
128 Lys Tyr Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp
129             195                 200                 205
131 Gln Ala Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp His Leu
132         210                 215                 220
134 Val Pro Arg Gly Ser Met Leu Arg Leu Leu Lys Arg Asn Ile Gln Ile
135 225                 230                 235                 240
137 Ser Lys Arg Ile Val Phe Thr Ile Leu Lys Gln Lys Ala Phe Lys Gly
138                 245                 250                 255
140 Asn His Pro Cys Val Pro Ser Val Cys Thr Ile Thr Tyr Ser Arg Phe
141             260                 265                 270
143 His Cys Leu Pro Asp Thr Leu Lys Ser Leu Leu Pro Met Ser Ser Lys
144         275                 280                 285
146 Thr Thr Leu Ser Met Leu Pro Gln Val Asn Ile Gly Ala Asn Ser Phe
147     290                 295                 300
149 Ser Ala Glu Thr Pro Val Asp Leu Lys Lys Glu Asn Glu Thr Glu Leu
150 305                 310                 315                 320
152 Ala Asn Ile Ser Gly Pro His Lys Lys Ser Thr Ser Thr Ser Thr Arg
153             325                 330                 335
155 Lys Arg Ala Arg Ser Ser Lys Lys Lys Ala Thr Asp Ser Val Ser Asp
156             340                 345                 350
158 Lys Ile Asp Glu Ser Val Ala Ser Tyr Asp Ser Ser Thr His Leu Arg
159         355                 360                 365
161 Arg Ser Ser Arg Ser Lys Lys Pro Val Asn Tyr Asn Ser Ser Ser Glu
162     370                 375                 380
164 Ser Glu Ser Glu Glu Gln Ile Ser Lys Ala Thr Lys Lys Val Lys Gln
165 385                 390                 395                 400
167 Lys Glu Glu Glu Glu Tyr Val Glu Glu Val Asp Glu Lys Ser Leu Lys
168             405                 410                 415
170 Asn Glu Ser Ser Ser Asp Glu Phe Glu Pro Val Val Pro Glu Gln Leu
171             420                 425                 430
173 Glu Thr Pro Ile Ser Lys Arg Arg Arg Ser Arg Ser Ser Ala Lys Asn
174         435                 440                 445
176 Leu Glu Lys Glu Ser Thr Met Asn Leu Asp Asp His Ala Pro Arg Glu
177     450                 455                 460
179 Met Phe Asp Cys Leu Asp Lys Pro Ile Pro Trp Arg Gly Arg Leu Gly
180 465                 470                 475                 480
182 Tyr Ala Cys Leu Asn Thr Ile Leu Arg Ser Met Lys Glu Arg Val Phe
183             485                 490                 495
185 Cys Ser Arg Thr Cys Arg Ile Thr Thr Ile Gln Arg Asp Gly Leu Glu
186             500                 505                 510
188 Ser Val Lys Gln Leu Gly Thr Gln Asn Val Leu Asp Leu Ile Lys Leu
189         515                 520                 525
191 Val Glu Trp Asn His Asn Phe Gly Ile His Phe Met Arg Val Ser Ser
192     530                 535                 540
194 Asp Leu Phe Pro Phe Ala Ser His Ala Lys Tyr Gly Tyr Thr Leu Glu
195 545                 550                 555                 560
```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/724,296

DATE: 05/30/2001
TIME: 15:37:00

Input Set : **A:\25-98a.app**
Output Set: **C:\CRF3\05302001\I724296.raw**

```
197 Phe Ala Gln Ser His Leu Glu Glu Val Gly Lys Leu Ala Asn Lys Tyr
198             565                 570                 575
200 Asn His Arg Leu Thr Met His Pro Gly Gln Tyr Thr Gln Ile Ala Ser
201             580                 585                 590
203 Pro Arg Glu Val Val Val Asp Ser Ala Ile Arg Asp Leu Ala Tyr His
204             595                 600                 605
206 Asp Glu Ile Leu Ser Arg Met Lys Leu Asn Glu Gln Leu Asn Lys Asp
207     610                 615                 620
209 Ala Val Leu Ile Ile His Leu Gly Gly Thr Phe Glu Gly Lys Lys Glu
210 625                 630                 635                 640
212 Thr Leu Asp Arg Phe Arg Lys Asn Tyr Gln Arg Leu Ser Asp Ser Val
213             645                 650                 655
215 Lys Ala Arg Leu Val Leu Glu Asn Asp Asp Val Ser Trp Ser Val Gln
216             660                 665                 670
218 Asp Leu Leu Pro Leu Cys Gln Glu Leu Asn Ile Pro Leu Val Leu Asp
219             675                 680                 685
221 Trp His His His Asn Ile Val Pro Gly Thr Leu Arg Glu Gly Ser Leu
222     690                 695                 700
224 Asp Leu Met Pro Leu Ile Pro Thr Ile Arg Glu Thr Trp Thr Arg Lys
225 705                 710                 715                 720
227 Gly Ile Thr Gln Lys Gln His Tyr Ser Glu Ser Ala Asp Pro Thr Ala
228             725                 730                 735
230 Ile Ser Gly Met Lys Arg Arg Ala His Ser Asp Arg Val Phe Asp Phe
231             740                 745                 750
233 Pro Pro Cys Asp Pro Thr Met Asp Leu Met Ile Glu Ala Lys Glu Lys
234             755                 760                 765
236 Glu Gln Ala Val Phe Glu Leu Cys Arg Arg Tyr Glu Leu Gln Asn Pro
237     770                 775                 780
239 Pro Cys Pro Leu Glu Ile Met Gly Pro Glu Tyr Asp Gln Thr Arg Asp
240 785                 790                 795                 800
242 Gly Tyr Tyr Pro Pro Gly Ala Glu Lys Arg Leu Thr Ala Arg Lys Arg
243             805                 810                 815
245 Arg Ser Arg Lys Glu Glu Val Glu Glu Asp Glu Lys
246             820                 825
249 <210> SEQ ID NO: 3
250 <211> LENGTH: 1161
251 <212> TYPE: DNA
252 <213> ORGANISM: Schizosaccharomyces pombe
254 <220> FEATURE:
255 <221> NAME/KEY: misc_feature
256 <222> LOCATION: (1)..(1161)
257 <223> OTHER INFORMATION: DNA sequence encoding UVDE protein, truncated at
258       amino acid residue 228.
260 <400> SEQUENCE: 3
261 gatgatcatg ctccacgaga gatgtttgat tgtttggaca aacccatacc ctggcgagga 60
262 cgattggggt atgcttgttt gaatactatt ttaaggtcaa tgaaggagag ggttttttgt 120
263 tcacgcacct gccgaattac aaccattcaa cgtgatgggc tcgaaagtgt caagcagcta 180
264 ggtacgcaaa atgttttaga tttaatcaaa ttggttgagt ggaatcacaa ctttggcatt 240
265 cacttcatga gagtgagttc tgatttattt cctttcgcaa gccatgcaaa gtatggatat 300
```

```
266 acccttgaat ttgcacaatc tcatctcgag gaggtgggca agctggcaaa taaatataat 360
267 catcgattga ctatgcatcc tggtcagtac acccagatag cctctccacg agaagtcgta 420
268 gttgattcgg caatacgtga tttggcttat catgatgaaa ttctcagtcg tatgaagttg 480
269 aatgaacaat taaataaaga cgctgtttta attattcacc ttggtggtac ctttgaagga 540
270 aaaaagaaa cattggatag gtttcgtaaa aattatcaac gcttgtctga ttcggttaaa 600
271 gctcgtttag ttttagaaaa cgatgatgtt tcttggtcag ttcaagattt attaccttta 660
272 tgccaagaac ttaatattcc tctagttttg gattggcatc atcacaacat agtgccagga 720
273 acgcttcgtg aaggaagttt agatttaatg ccattaatcc caactattcg agaaacctgg 780
274 acaagaaagg gaattacaca gaagcaacat tactcagaat cggctgatcc aacggcgatt 840
275 tctgggatga aacgacgtgc tcactctgat agggtgtttg actttccacc gtgtgatcct 900
276 acaatggatc taatgataga agctaaggaa aaggaacagg ctgtatttga attgtgtaga 960
277 cgttatgagt tacaaaatcc accatgtcct cttgaaatta tggggcctga atacgatcaa 1020
278 actcgagatg gatattatcc gcccggagct gaaaagcgtt taactgcaag aaaaaggcgt 1080
279 agtagaaaag aagaagtaga agaggatgaa aaataaaaat ccgtcatact ttttgattta 1140
280 tggcataatt tagccatctc c                                          1161
282 <210> SEQ ID NO: 4
283 <211> LENGTH: 371
284 <212> TYPE: PRT
285 <213> ORGANISM: Schizosaccharomyces pombe
287 <220> FEATURE:
288 <221> NAME/KEY: VARIANT
289 <222> LOCATION: (1)..(371)
290 <223> OTHER INFORMATION: Truncated version of the UVDE protein.
292 <400> SEQUENCE: 4
293 Asp Asp His Ala Pro Arg Glu Met Phe Asp Cys Leu Asp Lys Pro Ile
294  1               5                  10                  15
296 Pro Trp Arg Gly Arg Leu Gly Tyr Ala Cys Leu Asn Thr Ile Leu Arg
297             20                  25                  30
299 Ser Met Lys Glu Arg Val Phe Cys Ser Arg Thr Cys Arg Ile Thr Thr
300          35                  40                  45
302 Ile Gln Arg Asp Gly Leu Glu Ser Val Lys Gln Leu Gly Thr Gln Asn
303       50                  55                  60
305 Val Leu Asp Leu Ile Lys Leu Val Glu Trp Asn His Asn Phe Gly Ile
306  65                  70                  75                  80
308 His Phe Met Arg Val Ser Ser Asp Leu Phe Pro Phe Ala Ser His Ala
309             85                  90                  95
311 Lys Tyr Gly Tyr Thr Leu Glu Phe Ala Gln Ser His Leu Glu Glu Val
312          100                 105                 110
314 Gly Lys Leu Ala Asn Lys Tyr Asn His Arg Leu Thr Met His Pro Gly
315          115                 120                 125
317 Gln Tyr Thr Gln Ile Ala Ser Pro Arg Glu Val Val Val Asp Ser Ala
318       130                 135                 140
320 Ile Arg Asp Leu Ala Tyr His Asp Glu Ile Leu Ser Arg Met Lys Leu
321 145                 150                 155                 160
323 Asn Glu Gln Leu Asn Lys Asp Ala Val Leu Ile Ile His Leu Gly Gly
324             165                 170                 175
326 Thr Phe Glu Gly Lys Lys Glu Thr Leu Asp Arg Phe Arg Lys Asn Tyr
327             180                 185                 190
329 Gln Arg Leu Ser Asp Ser Val Lys Ala Arg Leu Val Leu Glu Asn Asp
```

VERIFICATION SUMMARY
PATENT APPLICATION: **US/09/724,296**

DATE: 05/30/2001
TIME: 15:37:01

Input Set : **A:\25-98a.app**
Output Set: **C:\CRF3\05302001\I724296.raw**

```
L:788  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:21
L:806  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:22
L:824  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:23
L:842  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:24
L:860  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:25
L:878  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:26
L:898  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:27
L:917  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:28
L:936  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:29
L:955  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:30
L:974  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:31
L:993  M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:32
L:1012 M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:33
L:1031 M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:34
L:1050 M:341 W:  (46) "n" or "Xaa" used, for SEQ ID#:35
```